

{7,"","","<I>","</I>","<TABLE>"}
procedure pcwrap HLRT (page <i>P</i>) scan to the line 7 in <i>P</i> while the next occurrence of <i>li</i> in <i>P</i> occurs before the next occurrence of <i>t</i> scan in <i>P</i> to next occurrence of <i>li</i> ; save position as start of item attribute scan in <i>P</i> to next occurrence of <i>ri</i> ; save position as end of item attribute scan in <i>P</i> to next occurrence of <i>lp</i> ; save position as start of price attribute scan in <i>P</i> to next occurrence of <i>rp</i> ; save position as end of price attribute return extracted , (item,price),. . . . } pairs

FIGURE 1
(PRIOR ART)

new U.S. Appl. based on Appln. Nos. 60/236,574
& 60/299,360 for Online Intelligent Information...
Express Mail Label No. EL864389877US
Attorney Docket No.: 2102680-990101
Gray Cary et al. – GTS/415-836-2500
Applicants: Victor Hsieh Sheet 1 of 36

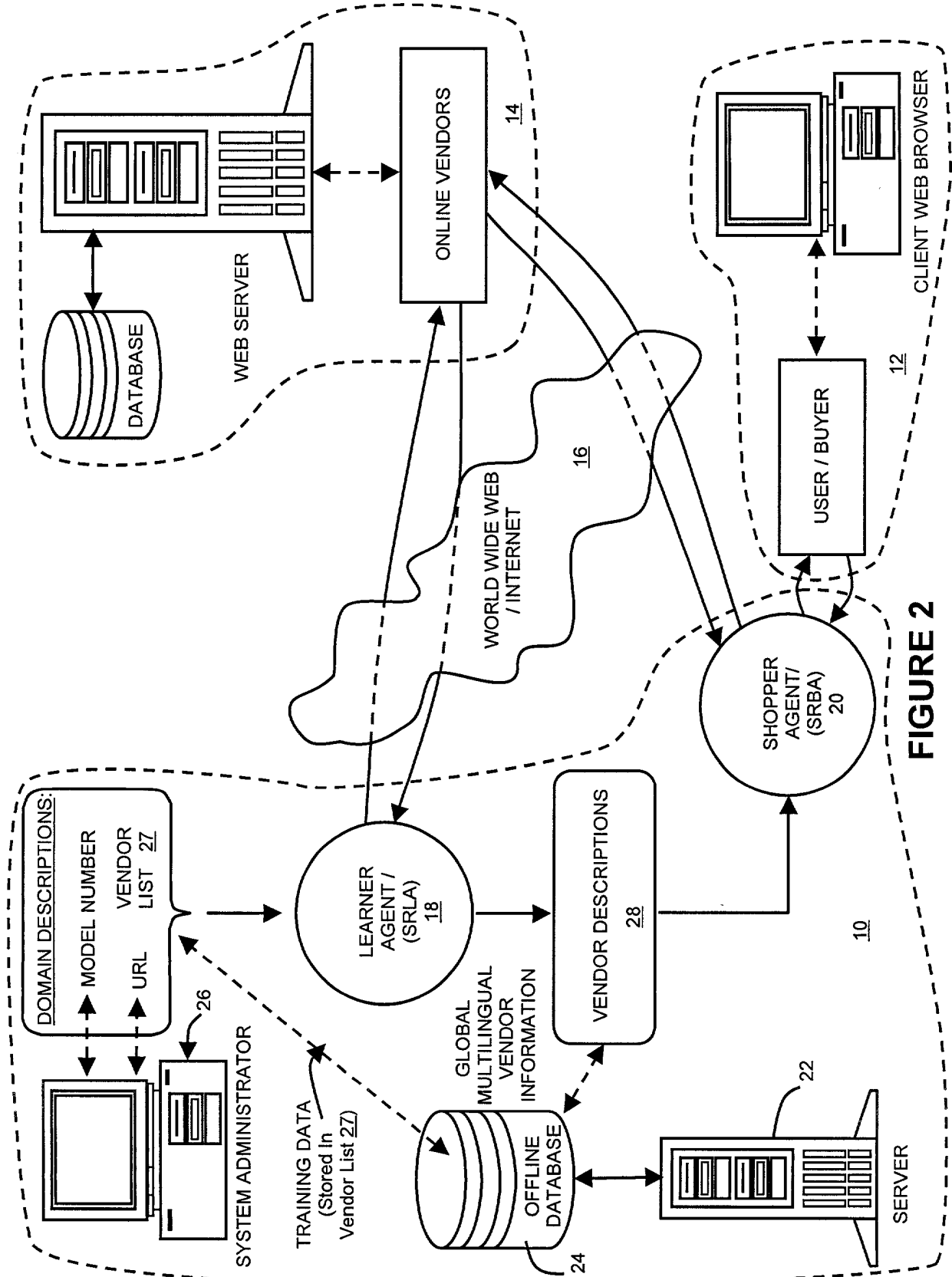


FIGURE 2

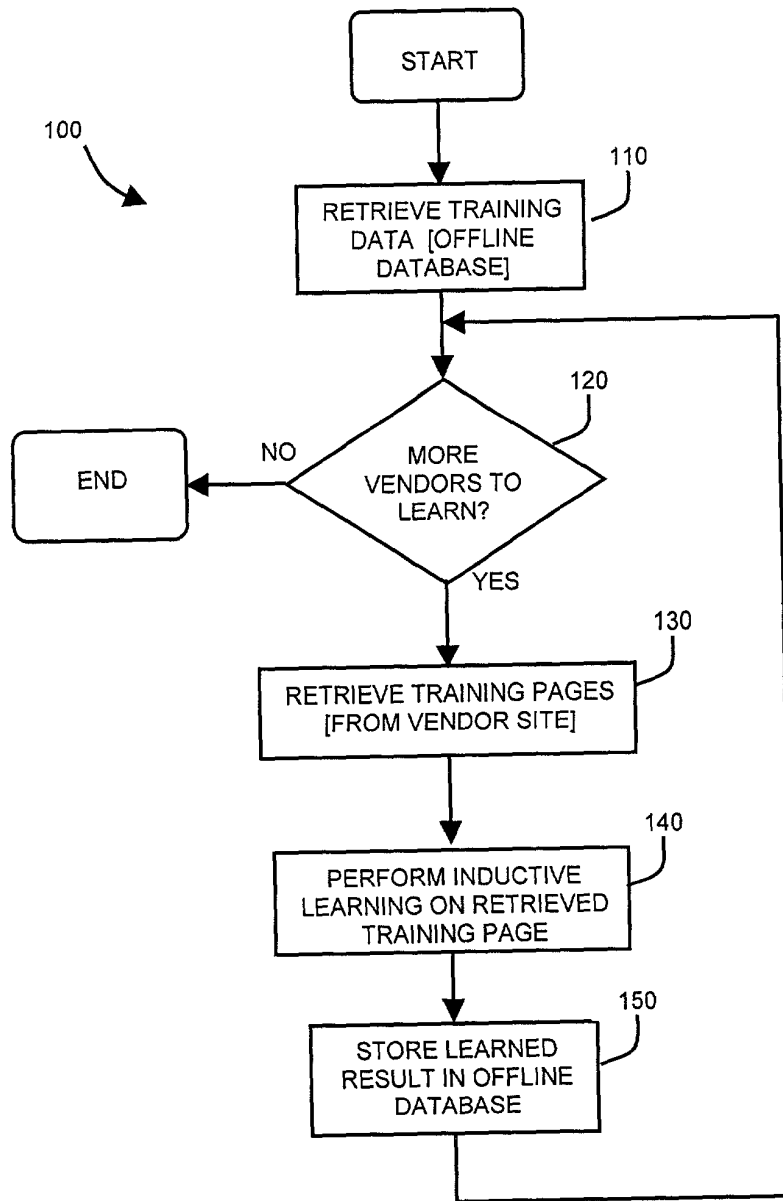


FIGURE 3

Element Name	Element Description
Vendor Name	Name of the online shop
Vendor URL	The URL of the online shop
Form URL	The URL to submit search data
Learning Domain	The domain used to learn the vendor description
Head	The end of header position of vendor's product pages. This element to learnt to reduce the product page searching space and thus shorten the product information extraction time
Tail	The start of footer position of vendor's product pages. Same as Head element, with this element the product information extraction time for Shopper Agent is shorten.
Left Delimiter of Item	The Shopper Agent uses this two delimiters to identify a product
Right Delimiter of Item	
Left Delimiter of Price	The Shopper Agent uses this two delimiters to locate a product's price information
Right Delimiter of Price	

FIGURE 4

EXAMPLE	
Element Name	Element Description
Vendor Name	800.com
Vendor URL	http://www.800.com
Form URL	http://www.800.com/search/srchrsIts.asp?qs=1&sIteentry=All&entry=
Learning Domain	Md
Head	1230
Tail	</BODY>
Left Delimiter of Item	/td width="12">
Right Delimiter of Item	</fon
Left Delimiter of Price	Your Price: \$
Right Delimiter of Price	

FIGURE 5

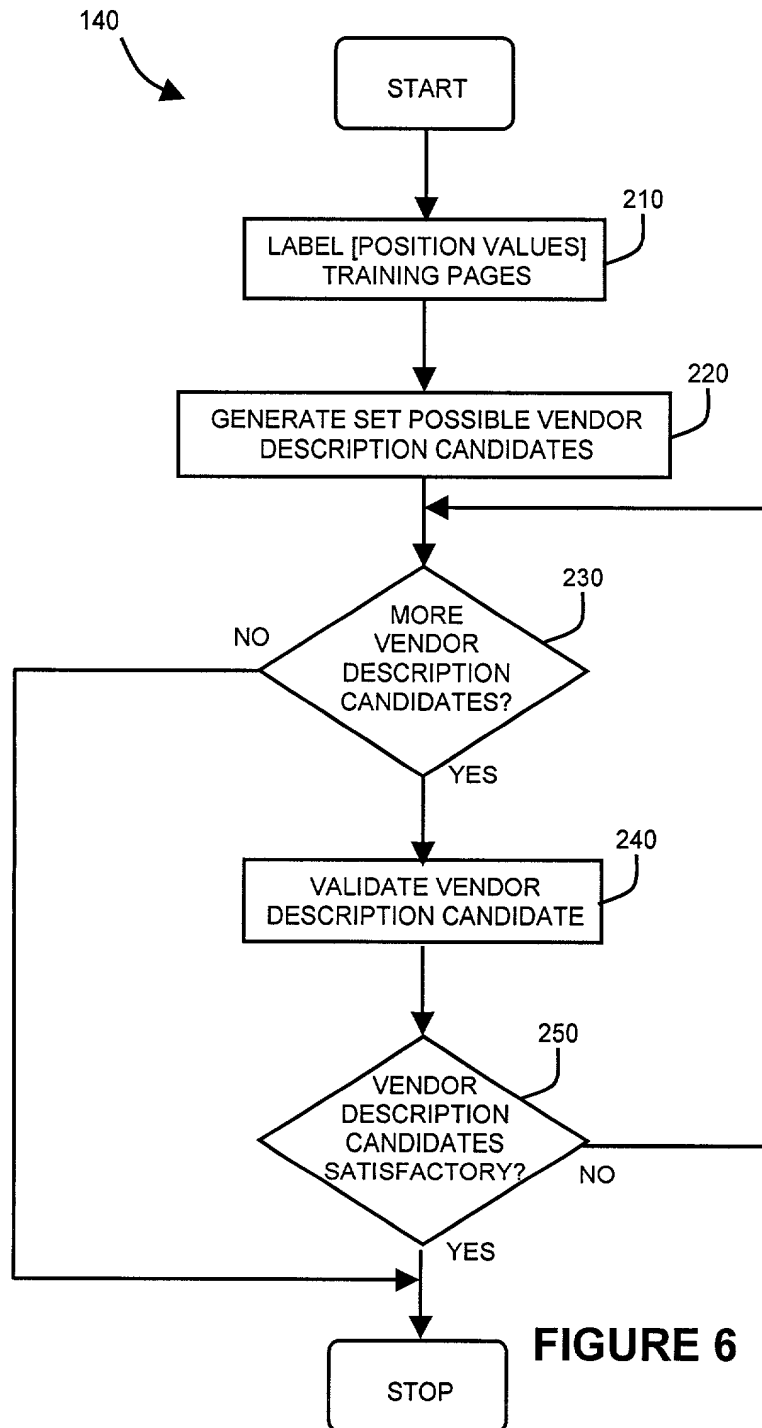


FIGURE 6

MD PRICE		<HTML> <TITLE>A Simple Product Catalogs</TITLE> <BODY> <H2> MD Price </H2> <TABLE BORDER=1> <TR BGCOLOR=ORANGE><TH> Model Number </TH> <TH> PRICE(US\$)</TH></TR> <TR><TD> HM381MD</TD><TD><I>399.95</I></TD></TR> <TR><TD> MD2070</TD><TD><I>599.95</I></TD></TR> <TR><TD> MD203</TD><TD><I>249.95</I></TD></TR> <TR><TD> MDR3</TD><TD><I>399.95</I></TD></TR> </TABLE> <P> <HR WIDTH=200 ALIGN=LEFT> <P> End Of The Product Catalog </BODY> </HTML>
Model Number	PRICE (US\$)	
HM381MD	399.95	
MD2070	599.95	
MD203	249.95	
MDR3	399.95	

FIGURE 7

Product Entry	Labels
{ HM381MD, 399.95 }	{ <<174, 180>, <197, 202>> }
{ MD2070, 599.95 }	{ <<229, 234>, <251, 256>> }
{ MD203, 249.95 }	{ <<283, 287>, <304, 309>> }
{ MDR3, 399.95 }	{ <<336, 339>, <356, 361>> }

FIGURE 8

Product Entry	Labels
{ PRODUCT, PRICE }	{ <PRODUCT LEFT DELIMITER, PRODUCT RIGHT DELIMITER>, <PRICE LEFT DELIMITER, PRICE RIGHT DELIMITER> }
***	***

FIGURE 9

Delimiter	Candidates
Item Left Delimiter	<TR><TD> TR><TD> R><TD> ><TD> <TD> TD> D> > B> >
Item Right Delimiter	</TD><TD>< > </TD><TD>< </TD><TD>< </TD><TD> </TD><TD </TD><T </TD>< </TD> </TD </T </ < </B </ <
Price Left Delimiter	</TD><TD>< > /B></TD><TD>< > B></TD><TD>< > ></TD><TD>< > </TD><TD>< > /TD><TD>< > TD><TD>< > D><TD>< > ><TD>< > <TD>< > TD>< > D>< > >< > < > > >

FIGURE 10A

FIGURE 10B

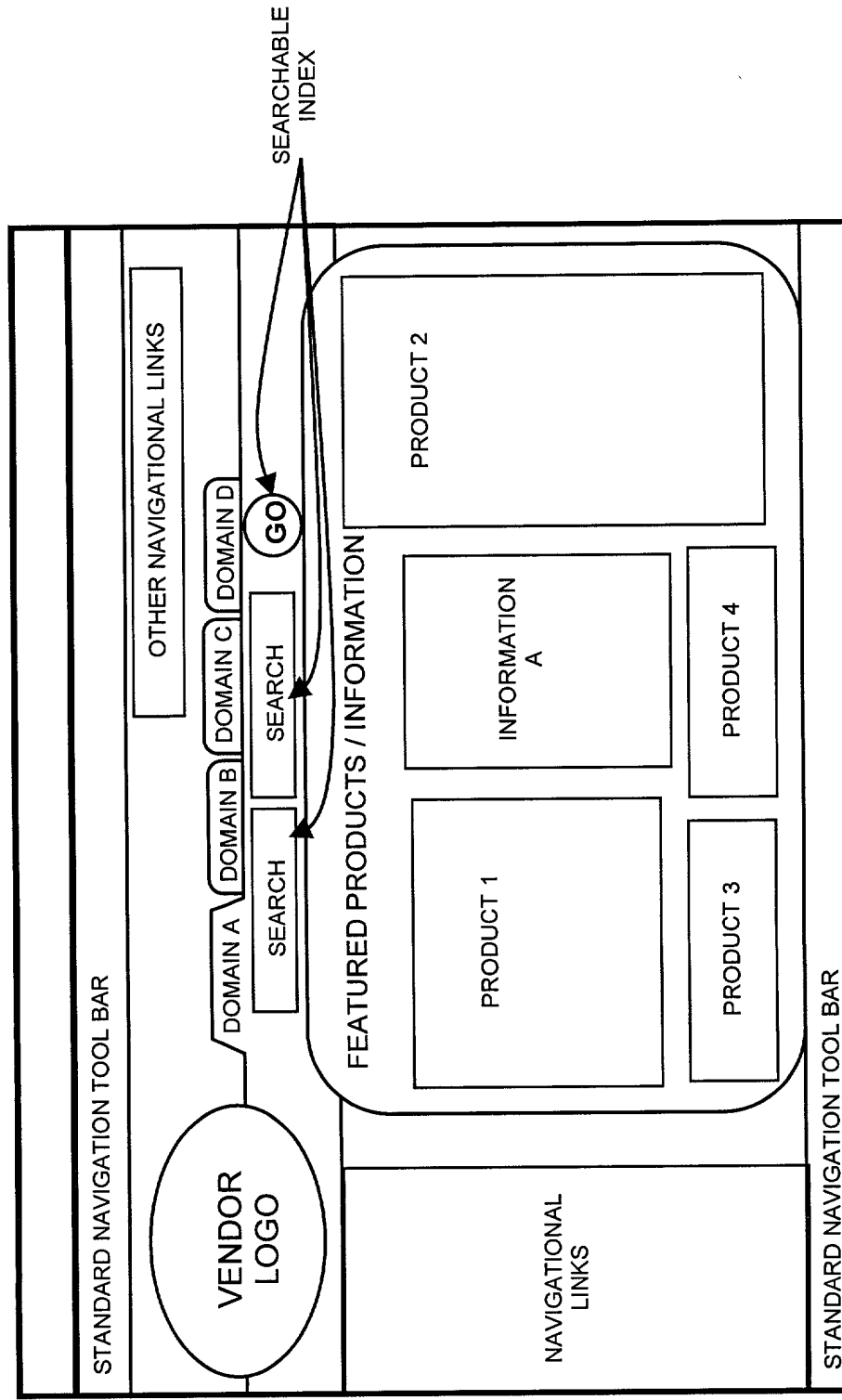


FIGURE 11

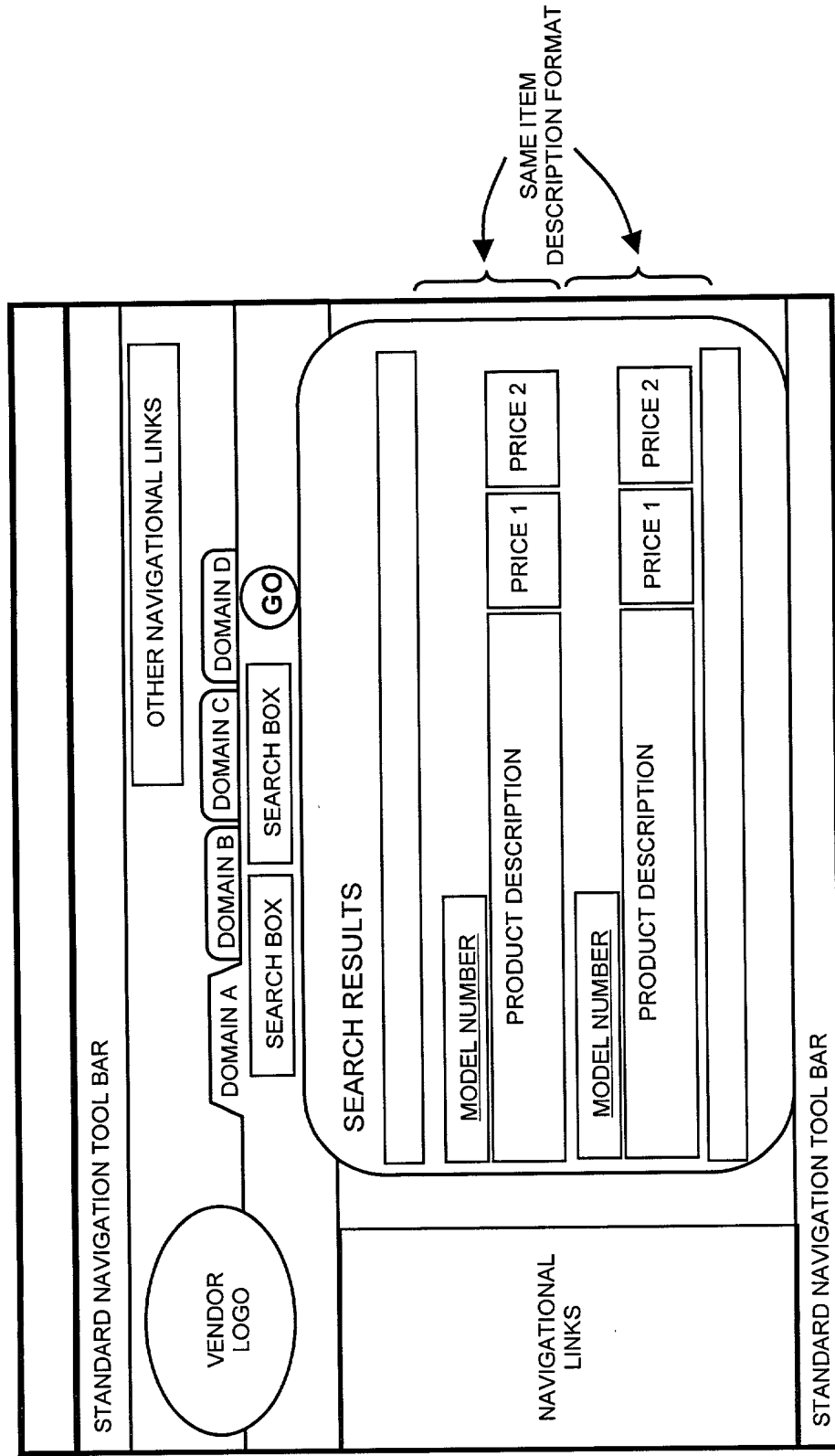


FIGURE 12

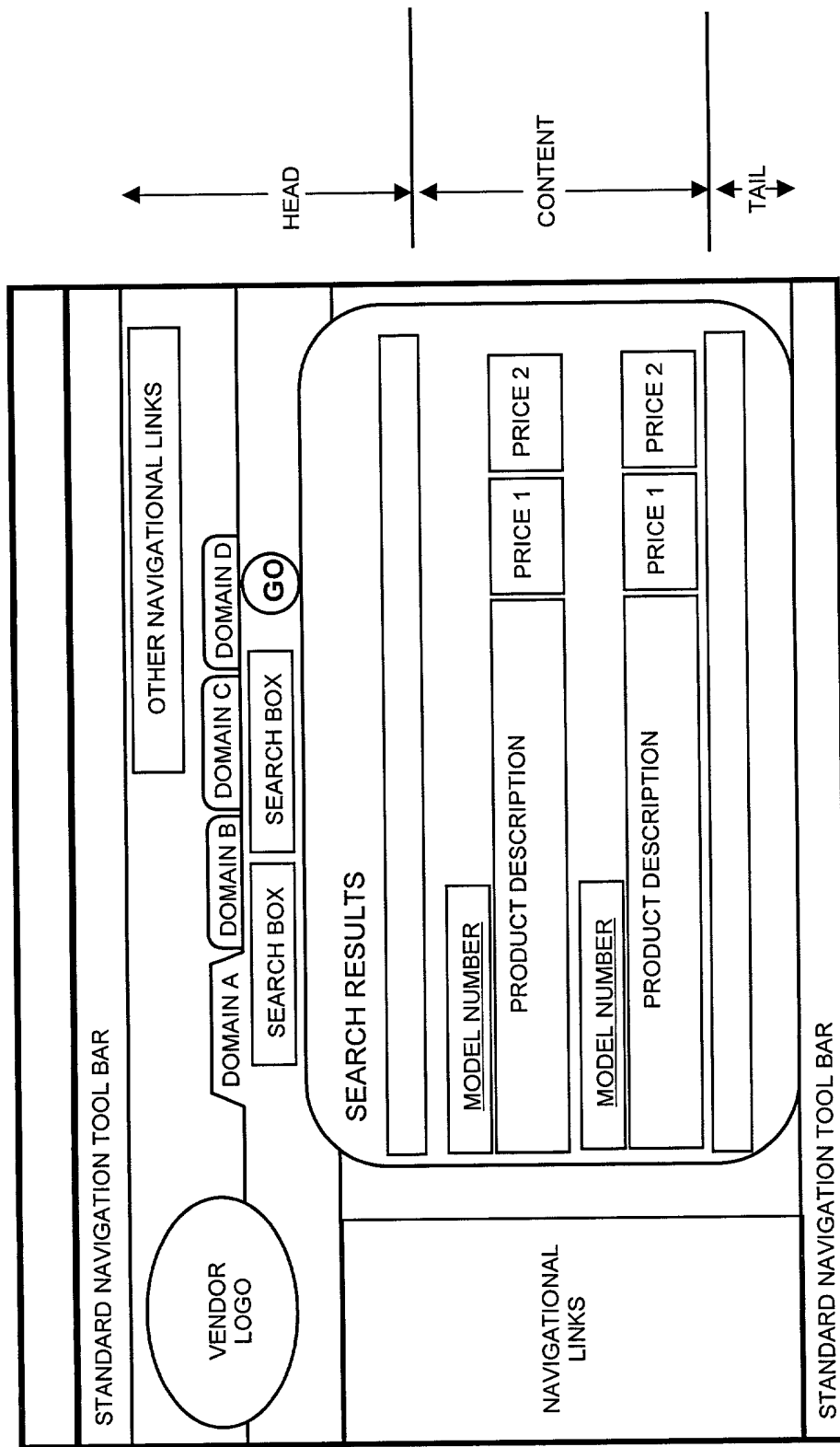


FIGURE 13

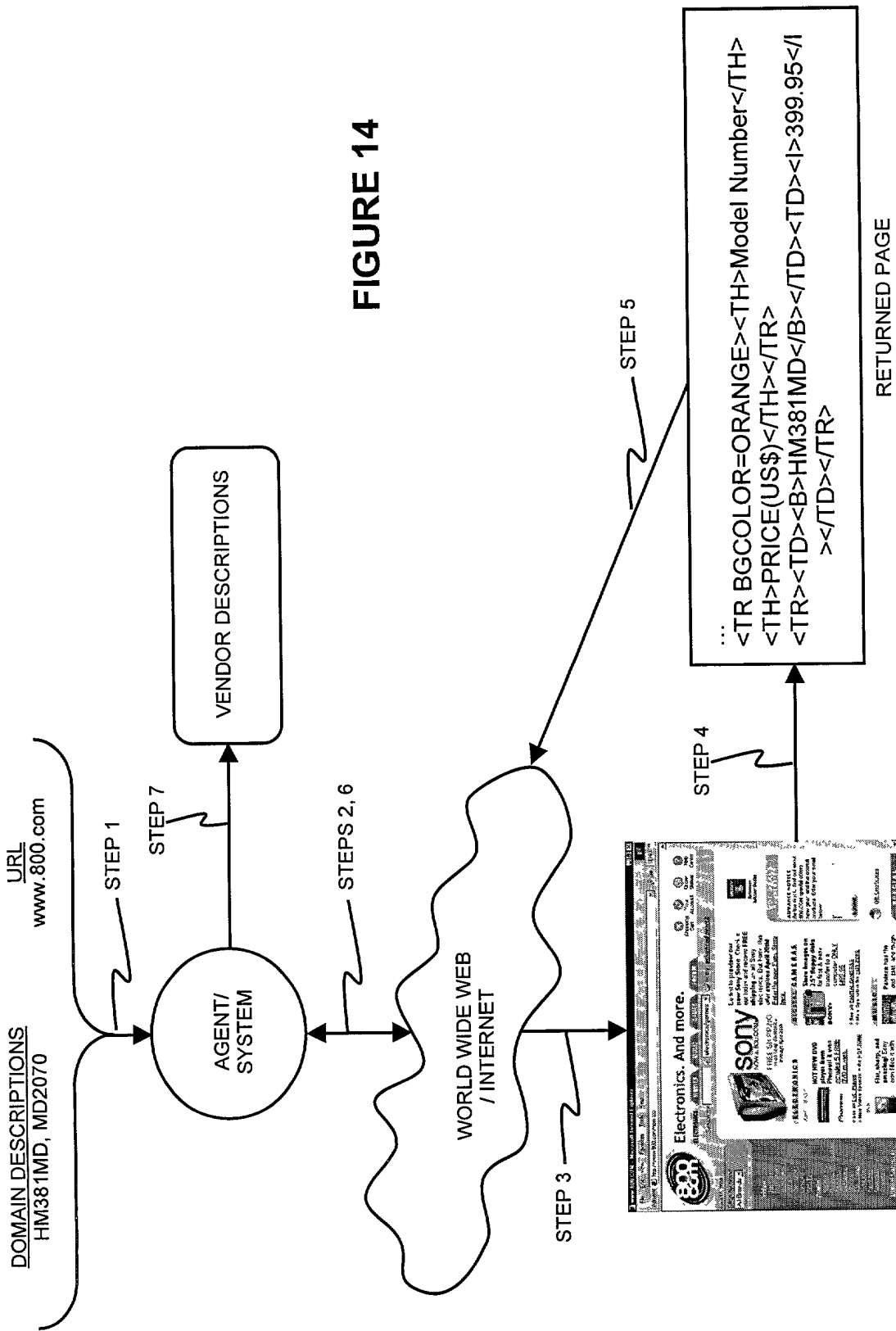


FIGURE 14

RETURNED PAGE

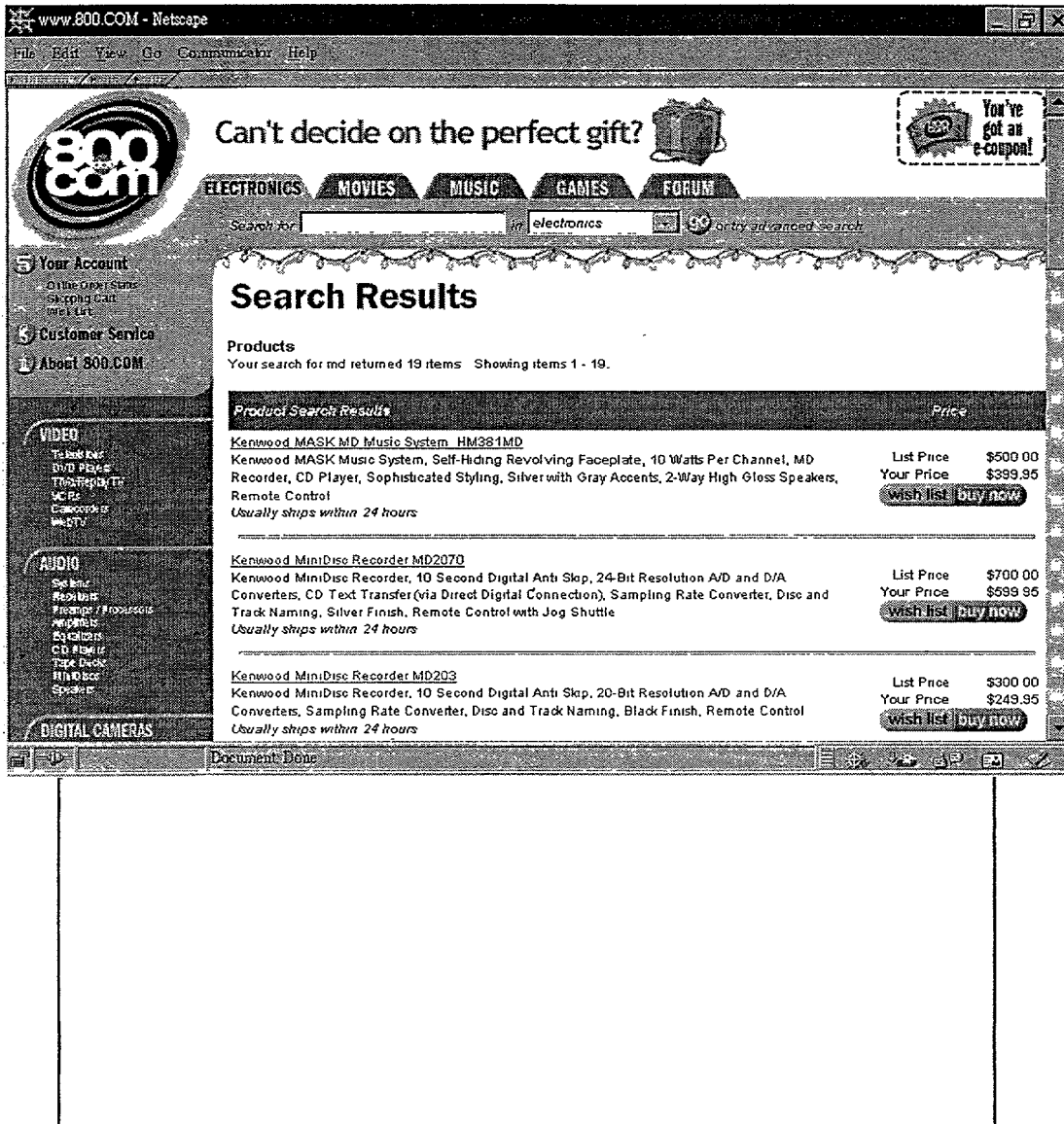


FIGURE 15A

60/299,360

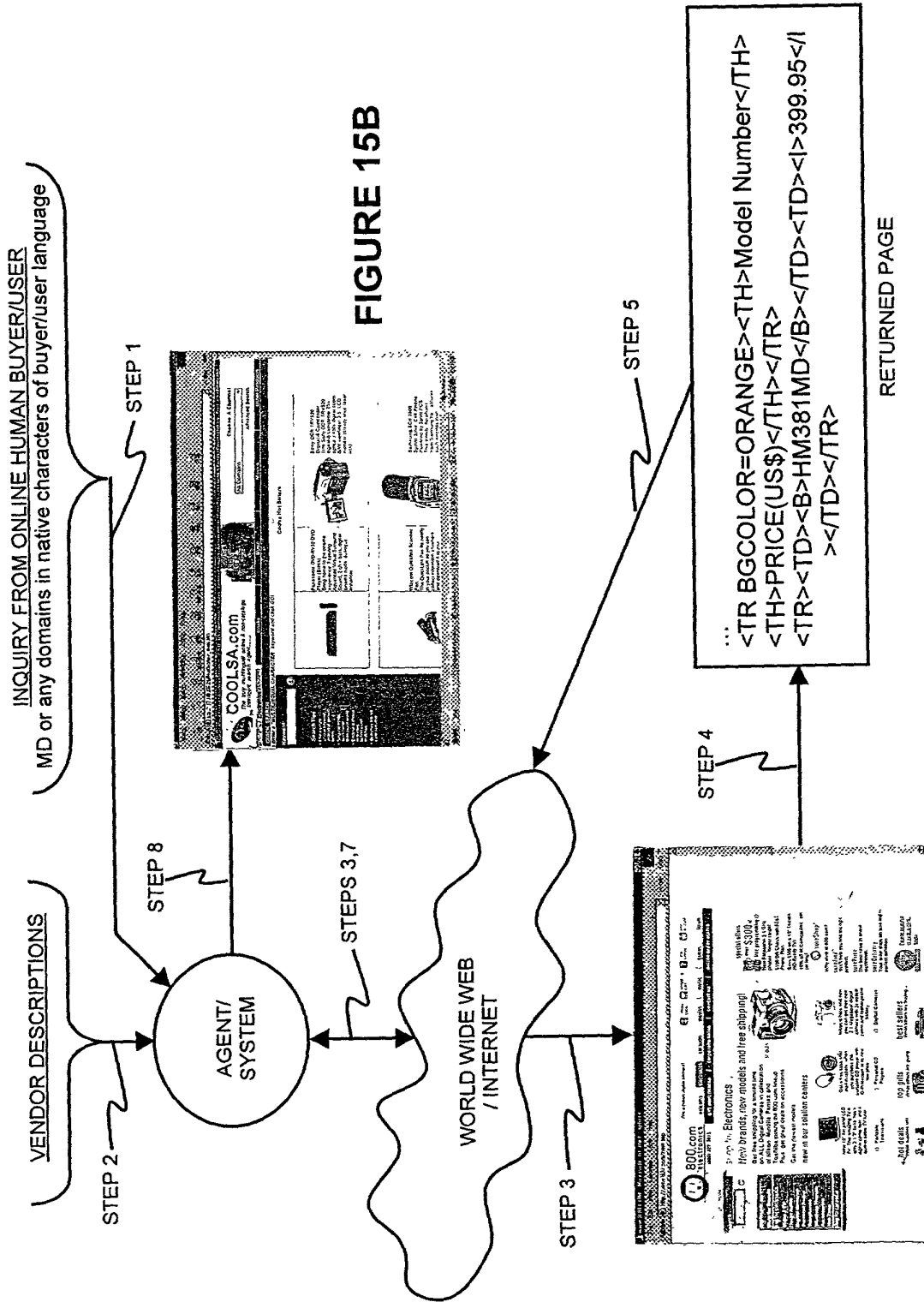


FIGURE 15B

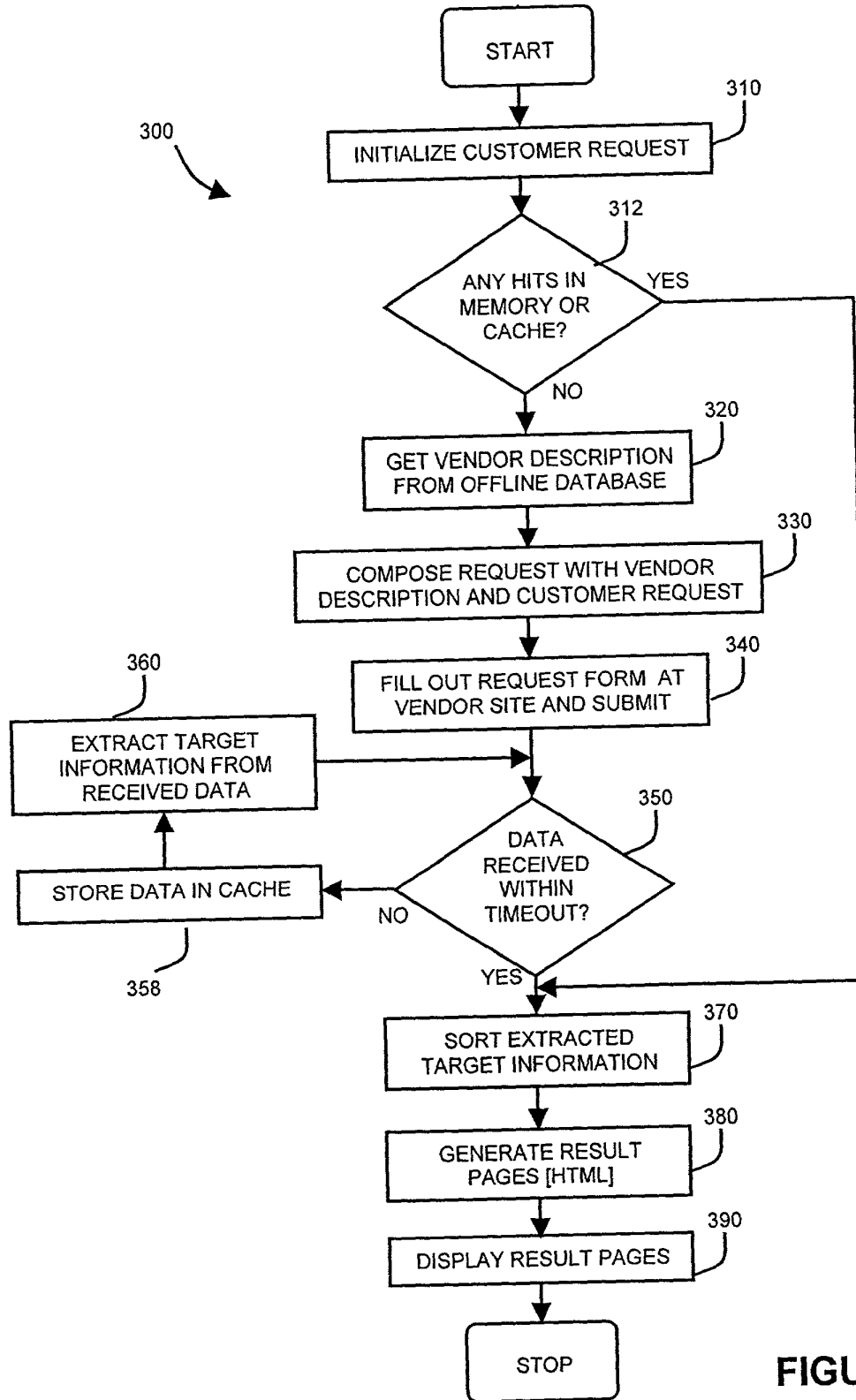


FIGURE 15C

Vendor Information [X]

Vendor Information | Add Vendor

Add Vendor

Details

☒ Provide Training Examples

Vendor Name:

URL:

Form's URL:

Learning Domain:

☐ Enter Wrapper For This Vendor

Head:

Tail:

Left Delimiter of Item:

Right Delimiter of Item:

Left Delimiter of Price:

Right Delimiter of Price:

Save

OK Cancel Apply

FIGURE 16

Vendor Information [X]

Vendor Information | Add Vendor

Add Vendor

Details

Vendor Name: ☒ Provide Training Examples

URL:

Form's URL:

Learning Domain:

☐ Enter Wrapper For This Vendor

Head	Tail
<input type="text" value="Left Delimiter of Item"/>	<input type="text" value="Right Delimiter of Item"/>
<input type="text" value="Left Delimiter of Price"/>	<input type="text" value="Right Delimiter of Price"/>

Save

OK Cancel Apply

FIGURE 17

Vendor Information

Vendor Information | Add Vendor

Vendor Information

Vendor Name: 1cache.com

Details

Vendor Name: 1cache.com

URL: http://www.1cache.com

Form's URL: http://st4.yahoo.com/cgi-bin/rssearch?catalog=1ca

Learning Domain: dvd

Wrapper

Head:

Tail:

Left Delimiter of Item:

Right Delimiter of Item:

Left Delimiter of Price:

Right Delimiter of Price:

FIGURE 18

Vendor Information

Vendor Information | Add Vendor

Vendor Information

Vendor Name: Search

Details:

Vendor Name:

URL:

Form's URL:

Learning Domain:

Wrapper:

Head: Tail:

Left Delimiter of Item: Right Delimiter of Item:

Left Delimiter of Price: Right Delimiter of Price:

Save

OK Cancel Apply

FIGURE 19

Training Data

Training Example

Vendor:

Vendor	Training Example
1cache.com	DVD Virtual Notebook Theater with I-Glas
1cache.com	JVC XV523GD Dolby Digital DVD Player
1cache.com	Pioneer DVL-919 - Combination LD/DVD

Vendor:
Example:

FIGURE 20

Learn Vendor Description

Learn Vendor Description

Learning Option

☒ Learn All
☐ Learn One

Vendor Name:

FIGURE 21

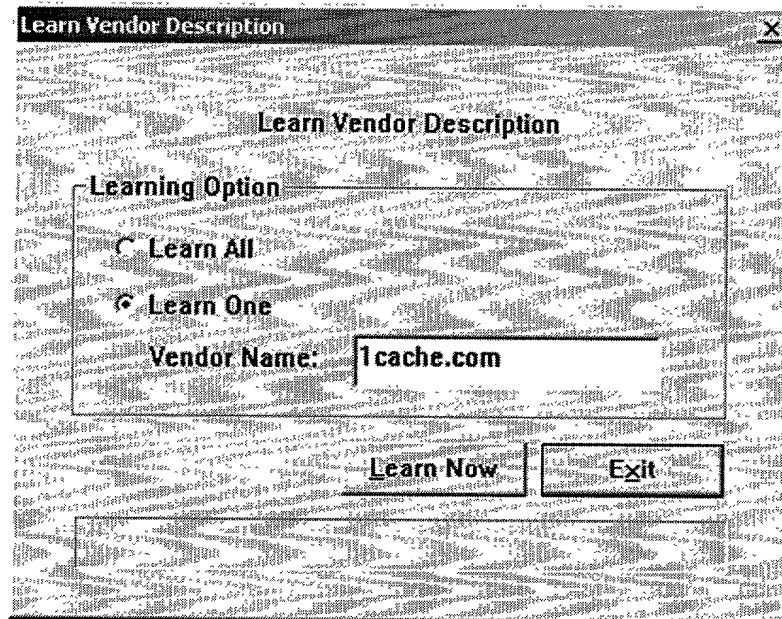


FIGURE 22

Vendor Information [X]

Vendor Information [Add Vendor]

Vendor Information

Vendor Name:

Details:

Vendor Name:

URL:

Form's URL:

Learning Domain:

Wrapper

Head: Tail:

Left Delimiter of Item: Right Delimiter of Item:

Left Delimiter of Price: Right Delimiter of Price:

FIGURE 23

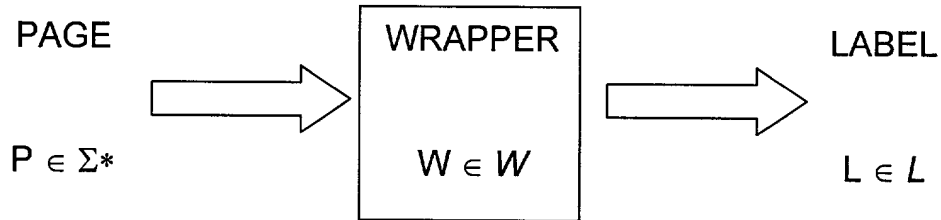


FIGURE 24

```

    procedure execHLRT (wrapper (h, li, ri, lp, rp, t), page P)
        m ← 0
        scan in P to line h
        while the next occurrence of li in P occurs before the next occurrence of t
            m ← m + 1
            scan in P to the next occurrence of li in P; save position as bm,i
            scan in P the next occurrence of ri; save position as em,i
            scan in P to the next occurrence of lp in P; save position as bm,p
            scan in P to next occurrence of rp; save position as em,p
        return label { . . . , <<bmi, emi>, <bmp, emp>> }
    
```

FIGURE 25

```

    Procedure learnHLRT (examples ε)
    1. Generate the candidate sets Candsr(l, ε), Candsl(l, ε), Candsr(p, ε), Candsl(p, ε).

    2. Enumerate the cross product of these candidate sets; each element W = (h, li, ri, lp, rp, t), of this cross product is a wrapper. Halt if W is satisfactory i.e., execLR(W, Pn) = Ln, for every (Pn, Ln) ∈ ε.
    
```

FIGURE 26

```

/*
* The main problem
*/
procedure learnHLRT (examples  $\epsilon$ ) for each  $1 < k < K$ 
    for each  $u \in \text{cands}_l(p, \epsilon)$ : if  $\text{valid}_l(u, p, \epsilon)$  then  $l_p \leftarrow u$  and terminate this loop
    for each  $u \in \text{cands}_r(i, \epsilon)$ : if  $\text{valid}_r(u, i, \epsilon)$  then  $r_i \leftarrow u$  and terminate this loop
    for each  $u \in \text{cands}_p(p, \epsilon)$ : if  $\text{valid}_p(u, p, \epsilon)$  then  $r_p \leftarrow u$  and terminate this loop
    for each  $u_{li} \in \text{cands}_l(i, \epsilon)$ 
    for each  $u_h \in \text{cands}_h(\epsilon)$ 
        for each  $u_t \in \text{cands}_t(\epsilon)$ 
            if  $\text{valid}_{li,ht}(u_{li}, u_h, u_t, \epsilon)$  then
                 $l_i \leftarrow u_{li}$ ,  $h \leftarrow u_h$ ,  $t \leftarrow u_t$ , and terminate these three loops
    return HLRT wrapper ( $h, l_i, r_i, l_p, r_p, t$ )

/*
*Generate a set of candidates for left delimiter of price attribute
*/
procedure  $\text{cands}_l(\text{attribute price, examples } \epsilon)$ 
    return the set of all suffixes of the shortest string in  $\text{neighbors}_l(\text{price}, \epsilon)$ 

/*
*Generate a set of candidates for right delimiter of attribute a. Here a could be item or price
*/
procedure  $\text{cands}_r(\text{attribute a, examples } \epsilon)$ 
    return the set of all prefixes of the shortest string in  $\text{neighbors}_r(a, \epsilon)$ 

/*
* Generate a set of candidates for a page's head
*/
procedure  $\text{cands}_h(\text{examples } \epsilon)$ 
    return the set of all substrings of the shortest string in  $\text{heads}(\epsilon)$ 

/*
*Generate a set of candidates for a page's tail
*/
procedure  $\text{cands}_t(\text{example } \epsilon)$ 
    return the set of all substrings of the shortest string in  $\text{tails}(\epsilon)$ 

/*
*Determine whether a particular candidate for the left delimiter of price is valid
*This procedure applies constraints C3
*/
procedure  $\text{valid}(\text{candidate } u, \text{attribute price, examples } \epsilon)$ 
    for each  $s \in \text{neighbors}_l(\text{price}, \epsilon)$ : if  $u$  is not a proper suffix of  $s$  then return FALSE
    
```

Figure 27A


```

        for each  $s \in \text{tails}(\epsilon)$ : if  $u$  is a sub string of  $s$  then return FALSE
        return TRUE

/*
* Determine whether a particular candidate for the right delimiter of item or price is
* valid
* This procedure applies constraints C1 and C2
*/
    procedure validi(candidate  $u$ , attribute  $a$ , examples  $\epsilon$ )
    for each  $s \in \text{attrs}(a, \epsilon)$ : if  $u$  is a sub string of  $s$  then return FALSE
    for each  $s \in \text{neighbors}(a, \epsilon)$ : if  $u$  is not a prefix of  $s$  then return FALSE
    return TRUE

/*
* Determine whether a particular combination of candidate  $u_h$ ,  $u_t$ , and  $u_l$  for  $h$ ,  $t$ , and
*  $l$  respectively are satisfactory.
*/
    procedure validh,t,l(candidates  $u_h, u_t, u_l$ , example  $\epsilon$ )
    for each  $s \in \text{heads}(\epsilon)$ 
        if  $u_h$  is not a sub string of  $s$  then return FALSE
        if  $u_l$  is not a proper suffix of scan ( $s, u_h$ ) then return FALSE
        if  $u_t$  occurs before  $u_l$  in scan ( $s, u_h$ ), then return FALSE
    for each  $s \in \text{tails}(\epsilon)$ 
        if  $u_t$  is not a sub string of  $s$  then return FALSE
        if  $u_l$  occurs before  $u_t$  in  $s$  then return FALSE
    for each  $s \in \text{seps}(\epsilon)$ 
        if  $u_l$  is not a proper suffix of  $s$  then return FALSE
        if  $u_t$  occurs before  $u_l$  in  $s$  then return FALSE
    return TRUE

/*
* Return a set of containing all values of a attribute, either item( $i$ ) or price( $p$ ) in each
* example
*/
    procedure attrs(attribute  $a$ , examples  $\epsilon$ )
        return  $\bigcup_{(P_n, L_n) \in \epsilon} \{P_n[b_{m,a}, e_{m,a}] \mid \langle \dots, (b_{m,k}, e_{m,k}), \dots \rangle \in L_n\}$ 

/*
* Return all strings to the left of an attribute, whether these strings are in the heads
* or the bodies of the pages
*/
    procedure neighbors(attribute  $a$ , examples  $\epsilon$ )
        if  $a=i$  then return  $\text{seps}(i, \epsilon) \cup \text{heads}(\epsilon)$  else return  $\text{seps}(a, \epsilon)$ 

```

FIGURE 27B

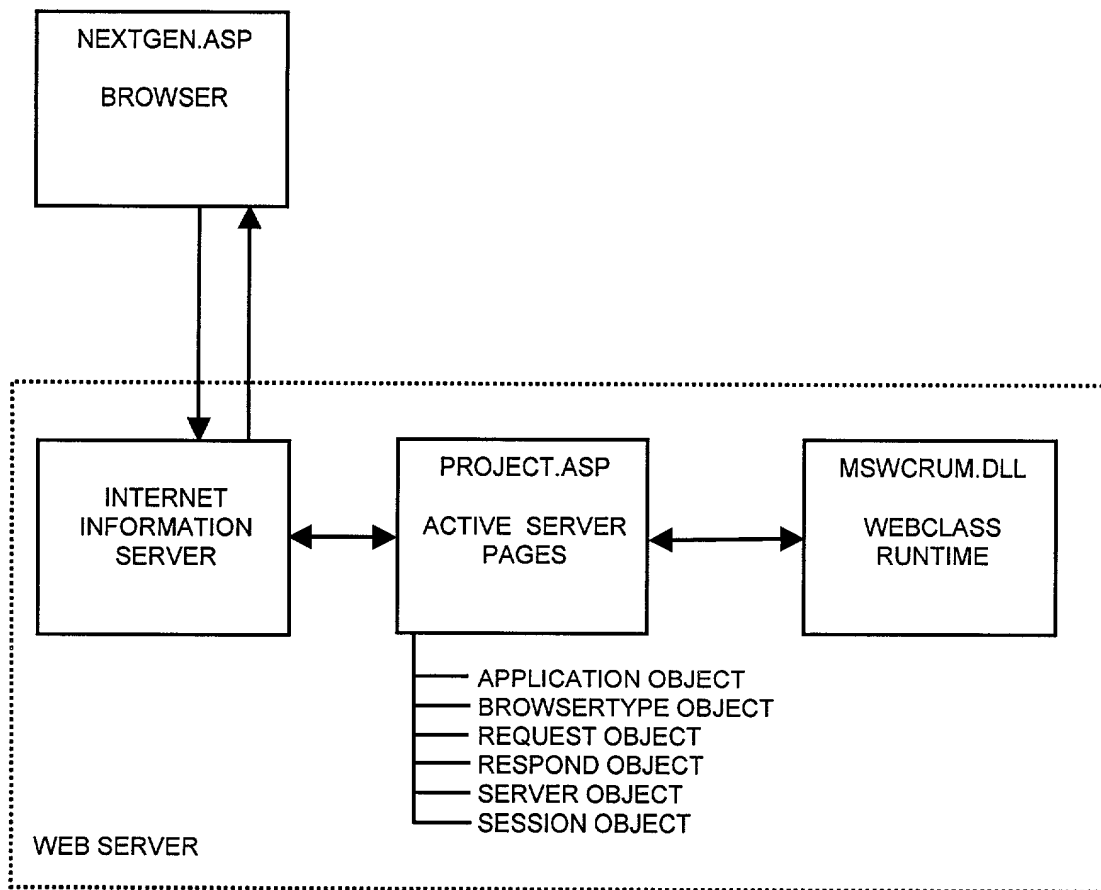


FIGURE 28

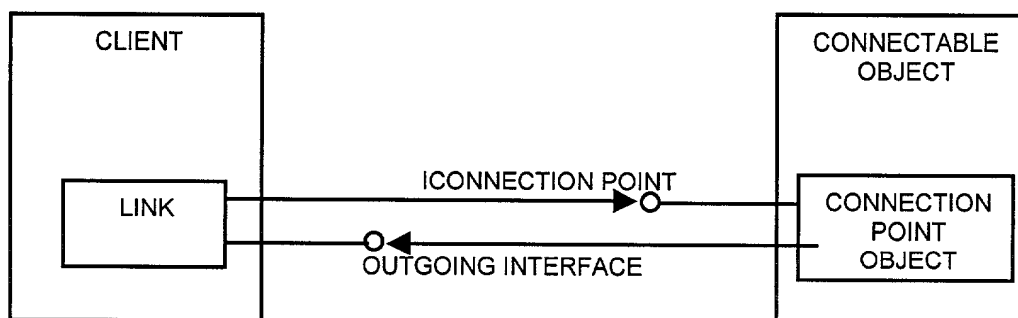


FIGURE 29

FIGURE 30

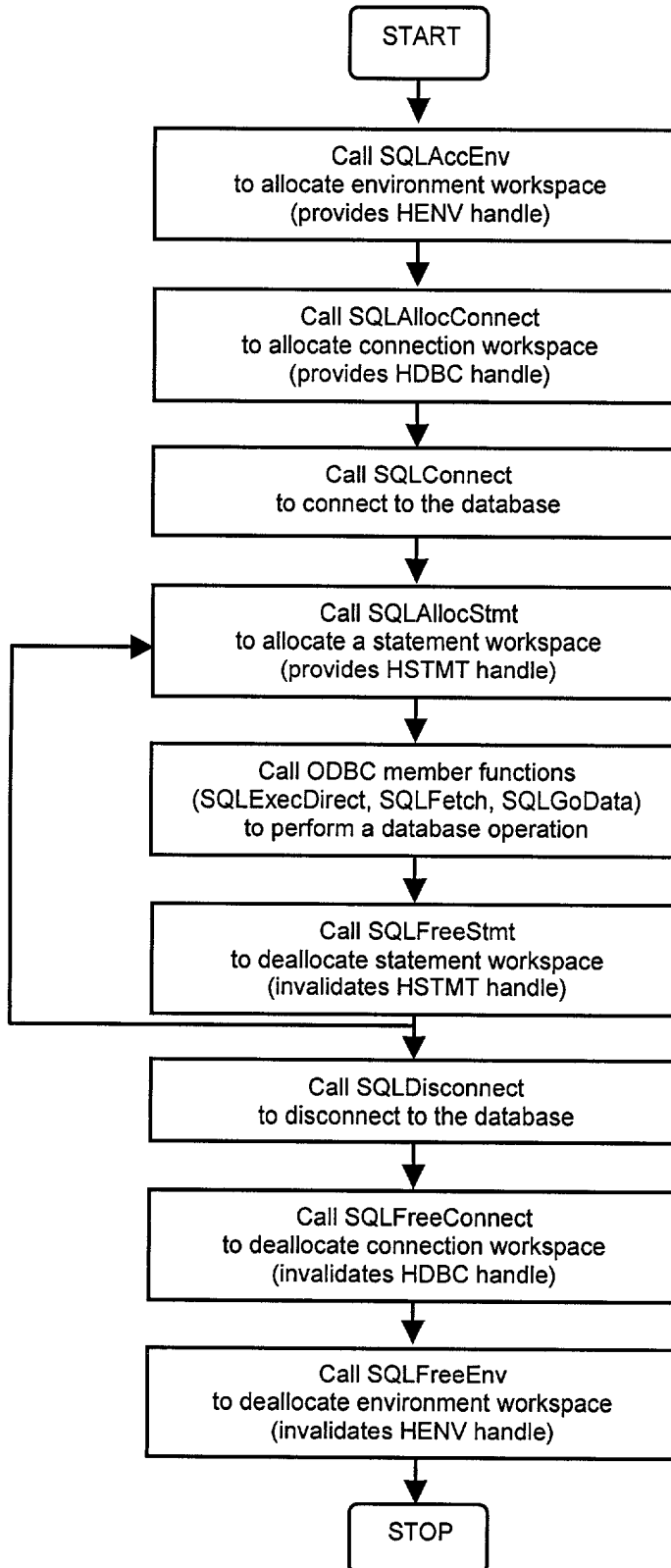
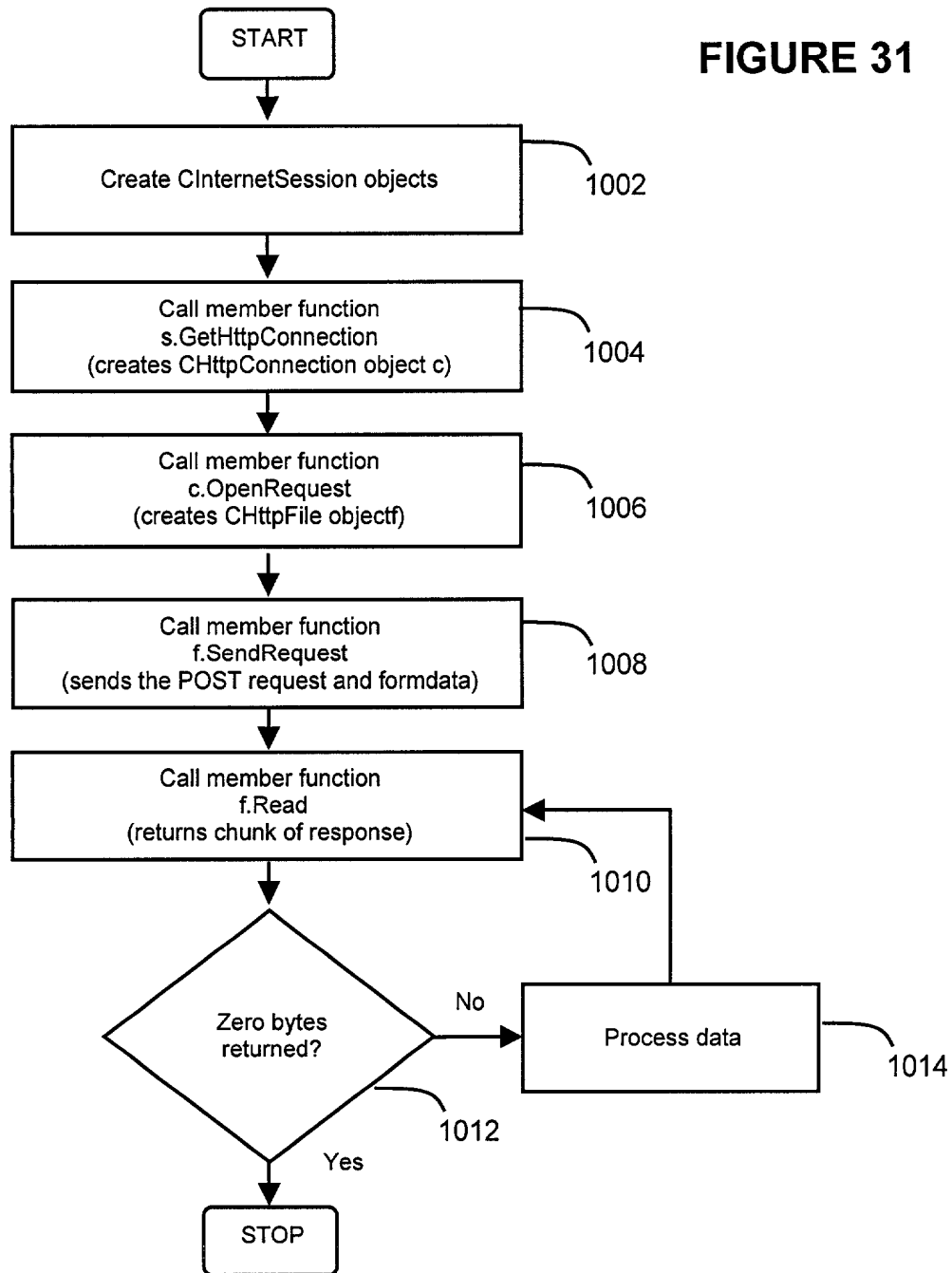


FIGURE 31



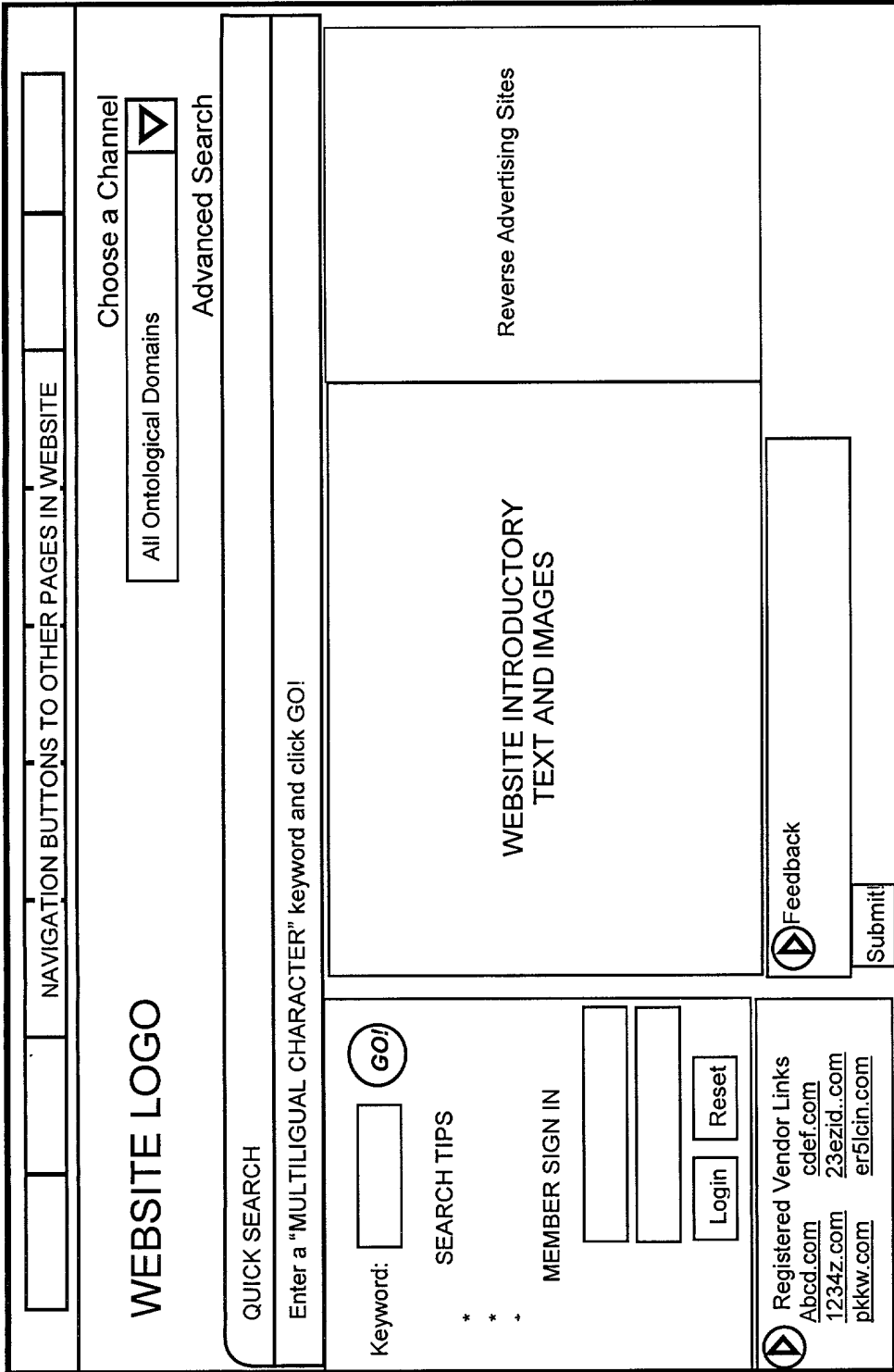


FIGURE 32

NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE									
<div>WEBSITE LOGO</div> <div>Choose a Channel</div> <div>ALL G2B PROCUREMENT DOMAINS ▾</div> <div>Advanced Search</div> <table><tr><td>DOMAIN A</td><td>DOMAIN B</td><td>DOMAIN C</td><td>DOMAIN D</td><td>OTHER DOMAINS</td></tr></table> <div>QUICK SEARCH</div> <div>Enter a "MULTILINGUAL CHARACTER" keyword and click GO!</div> <div><div>Keyword: <input type="text"/></div><div>GO!</div><div>SEARCH TIPS</div><div>* * *</div></div> <div><div>Company A Information about G2B Company A</div><div>Company G Information about G2B Company G</div><div>Reverse Advertising Sites</div></div> <div><div>Registered Vendor Links Abcd.com 1234z.com pkkw.com</div><div>Feedback</div><div>Submit!</div></div>					DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS
DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS					

FIGURE 33

NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE

WEBSITE LOGO

Choose a Channel

All G2B Tender Procurement Domains

Advanced Search

DOMAIN A

DOMAIN B

DOMAIN C

DOMAIN D

OTHER DOMAINS

ADVANCED SEARCH AGENTS ARE ON!

Enter a "MULTILINGUAL CHARACTER" keyword and click GO!

SEARCH TIPS

☐ Select All

Public Body

Speed

Air/Shipping/Ground/Free Overnight Del.

Ease of Navigation

☐ G2B Company G

⇄

EEEE

Time Out

Reserved Price Range

Manufacturer or OEM

Keyword:

From \$

To \$

GO!

FIGURE 34

<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="border: 1px solid black; width: 40px; height: 20px;"></div> <div style="border: 1px solid black; padding: 2px 5px;">NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE</div> <div style="border: 1px solid black; width: 40px; height: 20px;"></div> </div>							
<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 30%;"> <h2 style="margin: 0;">WEBSITE LOGO</h2> </div> <div style="width: 40%; text-align: right;"> <p>Choose a Channel</p> <div style="border: 1px solid black; padding: 2px; display: inline-block;">All Ontological Domains ▾</div> </div> </div>							
<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 30%;"> <p>Advanced Search</p> </div> </div>							
DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS			
<div style="border: 1px solid black; padding: 2px;">QUICK SEARCH</div>							
<p>Enter a "MULTILINGUAL CHARACTER" keyword and click GO!</p>							
<p>Keyword:</p> <div style="border: 1px solid black; width: 100px; height: 20px; margin: 5px 0;"></div> <p style="text-align: center; margin: 0;"> GO! </p> <p style="margin-top: 10px;">SEARCH TIPS</p> <p style="margin: 0;">* * *</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 30%; padding: 5px;"> <p>For B2B</p> <p>Trading Exchanges, Auctions, Vertical Portals, Production and Non-Production e-Procurement Supplies</p> </td> <td style="width: 30%; padding: 5px;"> <p>Agricultural / Farming Apparel</p> <p>Automotive/ Transportation Business Machines</p> <p>Computers & Electronics Industrial Equipment</p> <p>Medical & Scientific Equipment</p> </td> <td style="width: 40%; padding: 5px;"> <p>More Business to Business</p> <p>MRO/Raw Materials Office Furnishings Office Supplies Real Estate Restaurant/Hospitality Services</p> <p>Travel & Entertainment</p> </td> </tr> </table>				<p>For B2B</p> <p>Trading Exchanges, Auctions, Vertical Portals, Production and Non-Production e-Procurement Supplies</p>	<p>Agricultural / Farming Apparel</p> <p>Automotive/ Transportation Business Machines</p> <p>Computers & Electronics Industrial Equipment</p> <p>Medical & Scientific Equipment</p>	<p>More Business to Business</p> <p>MRO/Raw Materials Office Furnishings Office Supplies Real Estate Restaurant/Hospitality Services</p> <p>Travel & Entertainment</p>
<p>For B2B</p> <p>Trading Exchanges, Auctions, Vertical Portals, Production and Non-Production e-Procurement Supplies</p>	<p>Agricultural / Farming Apparel</p> <p>Automotive/ Transportation Business Machines</p> <p>Computers & Electronics Industrial Equipment</p> <p>Medical & Scientific Equipment</p>	<p>More Business to Business</p> <p>MRO/Raw Materials Office Furnishings Office Supplies Real Estate Restaurant/Hospitality Services</p> <p>Travel & Entertainment</p>					
<div style="display: flex; justify-content: space-between; align-items: center;"> <div style="width: 30%;"> <p>Registered Vendor Links</p> <p>Abcd.com 1234z.com pkkw.com</p> </div> <div style="width: 40%; text-align: right;"> <p>Feedback</p> <p>Submit</p> </div> </div>							

FIGURE 35

	NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE		
--	--	--	--

WEBSITE LOGO

Choose a Channel

All Ontological Domains ▼

Advanced Search

DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS
ADVANCED SEARCH AGENTS ARE ON!				

Enter a "MULTILINGUAL CHARACTER" keyword and click GO!

Select All	Multilingual B2B Suppliers			
Online Stores	Speed	Air/Shipping/Ground/Free Overnight Del.	Ease of Navigation	
<input type="checkbox"/> B2B Company 1	↑↑↑↑↑	⇒	EEEE	
<input type="checkbox"/> B2B Company 2	↑↑↑↑↑	⇒	EEEE	
<input type="checkbox"/> B2B Company 3	↑↑↑↑↑	⇒	EEEE	
<input type="checkbox"/> B2B Company 4	↑↑↑↑↑	⇒	EEEE	

SEARCH TIPS

* * *

Time Out

Price Range

From \$

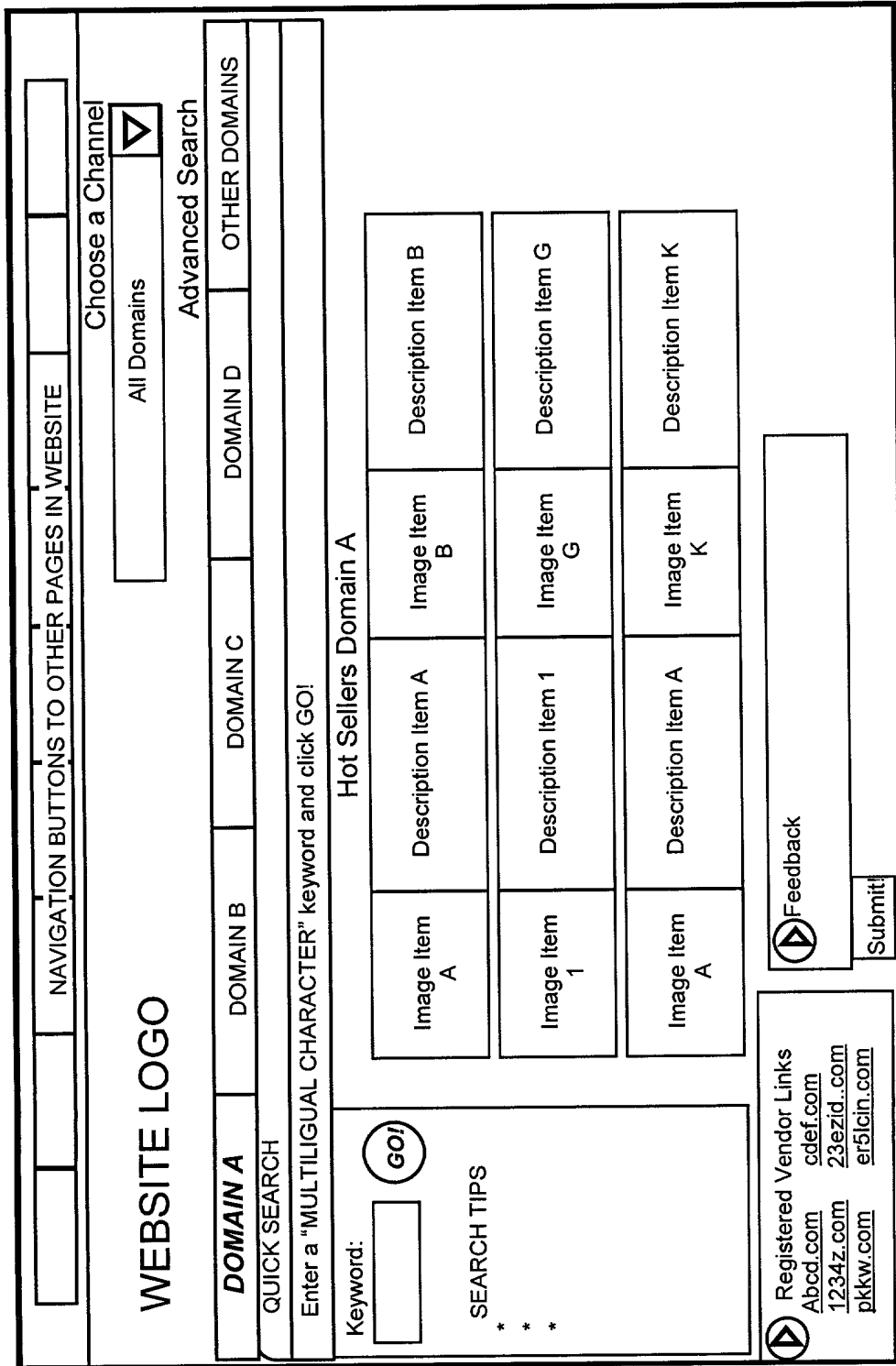
To \$

Manufacturer

Keyword:

GO!

FIGURE 36



NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE

WEBSITE LOGO

Choose a Channel

All Domains

Advanced Search

DOMAIN A

DOMAIN B

DOMAIN C

DOMAIN D

OTHER DOMAINS

ADVANCED SEARCH AGENTS ARE ON!

Enter a "MULTILINGUAL CHARACTER" keyword and click GO!

SEARCH TIPS

Select All

English Vendors

Online Stores

Speed

Air/Shipping/Ground/Free Overnight Del.

Ease of Navigation

☐ Store 1AAC

↑↑↑↑

⇒

EEEE

☐ Store 1443A

↑↑↑↑↑↑

⇒

EE

☐ Store 1443A

↑↑↑↑↑↑

⇒

EE

Select All

Chinese Vendors

☐ Store C5A3

↑↑↑↑↑↑

⇒

EEEE

☐ Store DRGWS4

↑↑↑↑↑↑

⇒

EEEE

Time Out

From \$

To \$

Price Range

Manufacturer

Keyword:

GO!

FIGURE 38

NAVIGATION BUTTONS TO OTHER PAGES IN WEBSITE

WEBSITE LOGO

Choose a Channel

ALL DOMAINS

▼

Advanced Search

DOMAIN A	DOMAIN B	DOMAIN C	DOMAIN D	OTHER DOMAINS
QUICK SEARCH				
Enter a "MULTILINGUAL CHARACTER" keyword and click GO!				

Keyword:

GO!

SEARCH TIPS

Search Results

Products

Your search for XX returned 5 Domain C items.

Model / Features	Store	Price
AMS7 XX Widget [Features]	Store 1443A	\$399.95 Buy now
AMS2 XX Widget [Features]	Store 1443A	\$329.95 Buy now
AMS7 XX Widget [Features]	Store DRGWS4	\$459.95 Buy now
AMS7 XX Widget [Features]	Store 4U\$U\$	\$499.95 Buy now
ZRTQ XX Widget [Features]	Store OOHE	\$529.95 Buy now

FIGURE 39